

Deep Collaborative Conjunctive Recommender Model based on Review Rating Prediction

Ananya Prakash Thakur¹, Shivangi Kaira², Mr.M Karthikeyan³

^{1&2} Student, Department of Computer Science and Engineering

³Assistant Professor, Department of Computer Science and Engineering
SRM Institute of Science and Technology Kattankulathur

ABSTRACT

Ratings or Recommendations provided by the users are a valid note & a significant reference to permissible users. However, these ratings are most probably biased or non-predictable in nature. This paper provides a Machine Learning based model to retrieve, analyze and predict a model to implement the user ratings and suggest a recommendation systems based on the user interests and necessity. Rating or recommendation models deploying the latest algorithms are a necessity on the recent times, wherein the markets rely undoubtedly on the reviews of the experienced users and customers. Significant issue rises in the classification of the dataset and analyzing and recommending the same. This paper implements the same based on the movie recommendation

Keywords

Rating Prediction, Machine / Deep Learning, Recommendation Model, Sentiment Analysis, Natural Language Processing

Introduction

User Opinion form an integral part of the reference models for any businesses. User Reviews and recommendations are a reliable source of information to the customers, which passes on the credibility of the reviews. With the ever increasing use of Smart phones and Internet, the risk of biased reviews is a great concern. More the extensive use of Mobile applications or services over the internet, more is the threat to vulnerability and credibility of the reviews.

Textual review based recommendation has gained enormous markets in the recent times. With the people sharing huge data over the internet and proving reviews of the services attended or purchased, this has become a capital market to those who wish to acquire the service. Recommendation sites have gained ultimatum in the recent years and gaining momentum still. Companies irrespective of the Tier and service, continuously render to the recommendation engines and American market shares more than 48% of the market share in the recommendation engines.

The growth rate of the recommendation engines over a period of ten years is enormous. On observation, from 35M to 104M reviews, the jump is so high and so is the demand to create a recommendation model to classify the reviews based on the ratings and provide nearest accurate recommendation systems to the users

Literature Review

Extensive works has been done earlier in the review based recommendation systems. Textual reviews are being processed in the existing systems. Machine Learning algorithms such as SVM Algorithm and Naïve Bayes Algorithms has already been deployed in the reviews based recommendation. However, these review classifications are only numeric based. Textual reviews are processed only on the basis of rating rather than sentimental analysis. The effectiveness of the model prediction is Non machine learning way of detecting based on characteristics,

properties,behavioral. Hence the disadvantages of the same may be Unawareness to detect and identify newly rated service, Non Machine learning approach & Not Reliable and not Realistic approach. Existing approaches covers mainly Direct HTTP or Java Script based reviews only. The data obtained is cleansed and is classified based on the rating only as Low rated and High Rated. Only High rated reviews are processed for the classification of the service, which rather tend to obstruct the whole concept of Review based recommendation system. The attributes then compared with the existing dataset and the rating is predicted and the recommendation model is provided. Thus the existing systems doesn't classify the textual ratings, neither classify them based on the bias.\

Topic

In this paper we have come up with idea to design and implement a review based recommendation systems for Movie recommendation for instance using Natural Language processing for textual reviews and biased / non biased reviews and provide a nearest accurate recommendation model.

First we have to gather dataset of reviews as training set. The data is then preprocessed to as to make it ready to apply the Algorithms and classifiers, so that is comparable and is classified. The dataset is then taken as reference model and the input data whichever received is then compared with the existing dataset and recommendation model generated.

With help machine learning algorithms such as SVM and Decision Tree algorithms we intend to make a comparison between training dataset and trained dataset. Support vector machine algorithms act as a classifier which is used to classify the reviews as biased or non-biased.

Improved accuracy in recommendation, Implementation of ML Algorithms such as SVM and Decision Tree, Computational analysis of the reviews using the Natural Language Processing techniques, less time consumption and accuracy results are the major advantages of the proposed system.

Datasets

The datasets for movie reviews are received as textual contents and ratings are in numeric. Datasets are collected from various streams of movie websites and the same is stored in a csv format for processing. Streaming big data is an analytic computing platform that is focused on accuracy and speed. The datasets have to be sampled so that is ensured that the data is not pertaining to one website or application. Multiple datasets from multiple sources pertain to multiple reviews and hence the accuracy of the recommendation model.Both the positive and negative reviews of data are taken into consideration. Data is preprocessed using the desired algorithms and is being read as csv files.

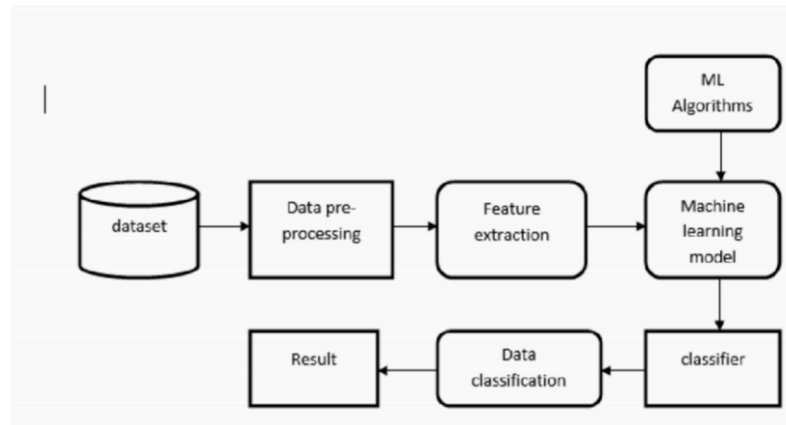


Figure 1. Flow Diagram of proposed system

Data preprocessing& feature extraction

The dataset collected is not structured i.e not available in the required format. It may contain multiple information that are not relevant to the processing of the data review. Large datasets require longer training time and because presence of stop words will reduce the recommendation accuracy. This is the reason why the data has to be text preprocessed. Preprocessing involves stemming, lowercase conversion, punctuation. Multiple NLP techniques are applied to the datasets so as to make it comparable to the existing dataset and thus generate a recommendation model which is near accurate.

Methods

Tokenization

Tokenization is applied to the chosen dataset. Tokenization is the process of making the content sentence into words, in other words as Tokens. This process of Tokenization is done to break the review sentence into the words, so that duplication of the words can be removed using TF –IDF

Term Frequency

Inverse Document Frequency is a widely used Natural Language Processing technique that is used to remove repetitive words from the selected dataset or review content. This process will remove the bag of words from the content and hence the processing speed and the computational algorithms and classifiers shall require minimum time for processing the content

Stemming The processed dataset is stemmed. The process of identifying the “stem” of the word is the idea behind this process. This will help to remove the frequency of the repetitive words and assist in processing of the data.

Data visualisation

Strong correlation between dataset categories and requested permissions, and introduce a method to visualize permissions usage in different categories. The aim of the work is to classify the provided reviews into datasets that are process able into several categories such as mentions and are categorically indexed to form a trained dataset that is comparable with any of the inputdata. This helps to classify whether the data review is malignant or genuine .

Methodology

The nature, sources and implications of sensitive reviews on content in review is found matches using the Decision Tree. They characterized and they have been proposed several approaches for dealing with security risks for enterprise. DWT is a method for recommendation based on three metrics, which evaluate: the occurrences of a specific subset of system calls, a weighted sum of a subset of permission that the application required, and a set of combinations of permissions.

Data Analysis

Machine learning techniques by analyzing the extracted features from the review data is processed using classifiers and a recommendation model created. Features used to classify are the presence of tags uses-permission and uses-feature into the manifest each of the review content. The recommendation model as in this case is illustrated as in Fig 3.4 wherein once the data input “batman” is entered, the system model would automatically compare the words to the existing data cluster and would suggest few movies based on the index, number of movies to show, no of review ratings of the movie and so on. The illustrative content and index can be segregated as per the requirement. Once, the Support Vector Machine model is trained then it is transferred to the learned model and added to the comparable dataset, so as to enhance the accuracy of the future predictions and make the recommendation model more accurate

Software

A. Python 2.5 / 3.5

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable Python is Object-Oriented – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

Packages such as numpy, matplotlib, pandas and Natural Language processing data packages are deployed in the process of implementing this paper.

B. Anaconda Navigator

Anaconda Navigator is the GUI that helps in visualizing the results in the form of graphics. Output predictions can be well visualized using the navigator

Discussions & Conclusion

Various results can be categorized from the total list of applications that has been tested. The trained dataset of applications are checked using a confusion matrix as listed below



Limitations and Future Studies

Different methodologies proposed by various researchers are considered, all of which show that Sentimental Analysis application using the Natural Language processing techniques and Machine Learning algorithms make the process of the Recommendation system model more accurate. From the performance criteria such as accuracy, clarity and specificity, different techniques have been recommended to increase the prognosis.

Acknowledgement

We will extend our work for various new algorithms for providing optimum results in context to existing techniques. Recommendation Model using the user review is always a complex and sensitive task, so preciseness and reliability will also plays an major role in the selection of the method

References

- [1] J. Horrigan, Online shopping, pew internet and american life proj-ect, Washington, DC, 2018, <http://www.pewinternet.org/Reports/2008/Online-Shopping/01SummaryofFindings.aspx> Online: accessed 8 Aug. 2014.
- [2] D. Pagano and W. Maalej, User feedback in the appstore: An em-pirical study, in Proc. IEEE Int. Requirements Eng. Conf. (Rio de Janeiro, Brazil), July 2013, pp. 125–134.
- [3] T. Chumwatana, Using sentiment analysis technique for analyzing Thai customer satisfaction from social media, 2015.
- [4] 4. T. Thiviya et al., Mobile apps' feature extraction based on user reviews using machine learning, 2019.
- [5] H. Hanyang et al., Studying the consistency of star ratings and reviews of popular free hybrid android and ios apps,

- [6] N. Kumari and S. Narayan Singh, Sentiment analysis on e-com-merce application by using opinion mining, in Proc. Int. Conf.-Cloud Syst. Big Data Eng. (Noida, India), Jan. 2016, pp. 320–325.
- [7] R. M. Duwairi and I. Qarqaz, Arabic sentiment analysis using su-pervised classification, in Proc. Int. Conf. Future Internet Things Cloud (Barcelona, Spain), Aug. 2014, pp. 579–583.
- [8] H. S. Le, T. V. Le, and T. V. Pham, Aspect analysis for opinion min-ing of vietnamese text, in Proc. Int. Conf. Adv. Comput. Applicat. (Ho Chi Minh, Vietnam), Nov. 2015, pp. 118–123.
- [9] H. Wang, L. Yue, and C. Zhai, Latent aspect rating analy-sis on review text data: A rating regression approach, in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining (Washington, D.C., USA), July 2010, pp. 783–792.
- [10] K. Dave, S. Lawrence, and D. M. Pennock, Mining the peanut gal-lery:Opinion extraction and semantic classification of product reviews, in Proc. Int. Conf. World Wide Web (New York, USA), 2003, pp. 519–528.
- [11] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: Sentiment classi-fication using machine learning techniques, in Proc. ACL-02 Conf. Empirical Methods Natural Language Process. (Stroudsburg, PA, USA), 2002, pp. 79–86.
- [12] C. Cardie et al., Combining low-level and summary representations of opinions for multi-perspective question answering, New direc-tions in question answering, 2003, pp. 20–27.
- [13] Takamura, T. Inui, and M. Okumura, Extracting semantic orientations of words using spin model, in Proc. Annu. Meeting Association Comput. Linguistics (Ann Arbor, MI, USA), 2005, pp. 133–140.
- [14] Buche, D. Chandak, and A. Zadgaonkar, Opinion mining and analysis: A survey, arXiv preprint arXiv:1307.3336, 2013.
- [15] M. Suleman, A. Malik, and S. S. Hussain, Google play store app ranking prediction using machine learning algorithm, Urdu News Headline, Text Classification by Using Different Machine Learning Algorithms, 2019.
- [16] F. Sarro et al., Customer rating reactions can be predicted purely using app features, in Proc. IEEE Int. Requirements Eng. Conf. (Banaf, Canada), Aug. 2018, pp. 76–87